Modeling Fine-grained Information via Knowledge-aware Hierarchical Graph for Zero-shot Entity Retrieval

Taiqiang Wu, Xingyu Bai, Weigang Guo, Weijie Liu, Siheng Li, Yujiu Yang

Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

Contribution Highlights

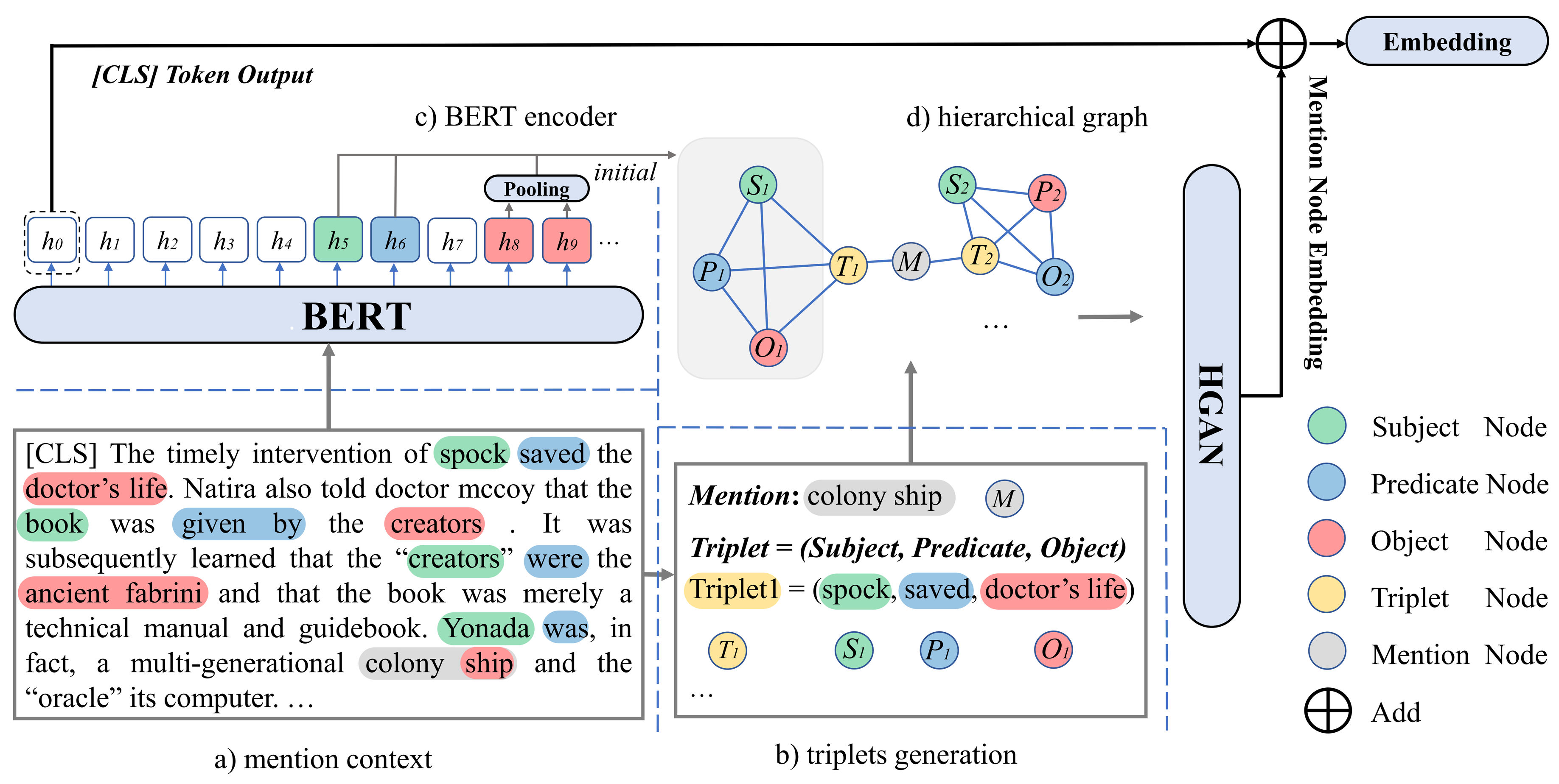


Figure: Overview of the mention encoder in our proposed GER.

In this work, we explore how to get comprehensive representations for mentions and entities in zero-shot entity linking task. The main contributions of our paper are as follows:

- We show that coarse-grained sentence embeddings can not fully model the mentions/entities, and more **fine-grained information** is necessary for the zero-shot entity retrieval task.
- We propose **GER**, which learns extra fine-grained information about mentions/entities as complementary to coarse-grained sentence embeddings. Particularly, we construct a mention/entity central graph and design a novel Hierarchical Graph Attention Network.
- We evaluate GER on several zero-shot entity retrieval datasets, and experimental results demonstrate that our framework achieves significant improvements compared with previous models.

Introduction

Background

- Entity Linking (EL) is a task of linking mentions in unstructured context to referent entries in a structured Knowledge Base. There are two assumptions under the zero-shot setting: (1) labeled mentions for the target domain are unavailable, (2) mentions and entities are only defined through textual descriptions (a.k.a. mention context and entity description).
- Most zero-shot entity linking systems follow a two-stage pipeline: Candidate Entity Retrieval, where top-k candidate entities are retrieved based on scores such as inner product values of the mention vector and entity vector, and Candidate Entity Ranking, where the candidates are ranked to find the most probable one. In this paper, we focus on **first-stage entity retrieval**, since the overall accuracy is capped by recall performance in this stage.

Previous Methods for Entity Retrieval

Given the mention contexts and entity descriptions in the source domain simultaneously, a general way is to embed them in a dense space and calculate similarity scores. BERT-based bi-encoder has been widely employed to represent the mentions/entities via the **sentence embeddings** (output of the [CLS] token) of corresponding mention contexts/entity descriptions.

Our Insights

- Intuitively, the sentence embeddings model the information of whole sentence rather than the mention/entity. When the attention scores from the [CLS] token to mentions/entities are relatively low, such sentence embeddings may be misled by other high-attention words, leading to a shift in the semantic vector space.
- Moreover, one intuitive idea is to adopt the output of mentions/entities from BERT rather than the [CLS] token. However, the output of mentions/entities is highly similar to the [CLS] token due to the over-smoothing problem. More information about mentions/entities besides BERT outputs is required.

Proposed Graph enhanced Entity Retrieval (GER)

Our key insight is to learn extra **fine-grained information about mentions/entities as complementary to coarse-grained sentence embeddings**.

- First, we employ the sequence prediction model to extract Subject- Predicate-Object (SPO) triplets as knowledge units.
- Second, we build a graph by connecting the knowledge units to the corresponding mention/entity. Such graph design allows mention/entity to aggregate information from these knowledge units.
- To avoid the graph bottleneck for the central mention/entity node, we construct a hierarchical graph to reduce neighboring nodes and then design a novel Hierarchical Graph Attention Network (HGAN).
- Finally, we employ the output of the central mention/entity node as fine-grained information, since they capture information at word level.

Experiments

Table: *Recall@K* (R@K) results on the test set of ZESHEL dataset, which is the average of 5 runs with different random seeds. * denotes for the results we reproduce. † denotes for the results taken from their papers. Best results are shown in bold.

Method	**R@1**	**R@4**	**R@8**	**R@16**	**R@32**	**R@50**	**R@64**
BM25 †	-	-	-	-	-	-	69.13
BLINK †	-	-	-	-	-	-	82.06
ARBORESCENCE	-	-	-	-	-	-	85.11
BLINK *	38.01	62.08	69.19	75.39	80.03	82.69	83.98
BERT Mean Pooling	33.65	57.74	65.17	71.38	75.85	78.66	80.14
BERT Max Pooling	36.94	60.42	68.34	73.83	78.40	81.09	82.65
BLINK + BERT Mean Pooling	34.12	58.41	66.19	72.24	76.93	79.79	81.16
BLINK + BERT Max Pooling	38.45	63.46	70.68	76.72	81.11	83.63	84.83
GER (ours)	**42.86**	**66.48**	**73.00**	**78.11**	**82.15**	**84.41**	**85.65**

→ Our proposed GER outperforms all baselines on the ZESHEL benchmark, which demonstrates the effectiveness of GER on getting more comprehensive representations for mentions/entities. Moreover, both fine-grained information and coarse-grained sentence embedding are crucial to represent mentions/entities.

Ablation Studies

Table: The ablation study results of the dual-encoder architecture. (BERT, BERT) is the baseline BLINK while (BERT+HGAN, BERT+HGAN) is our proposed GER.

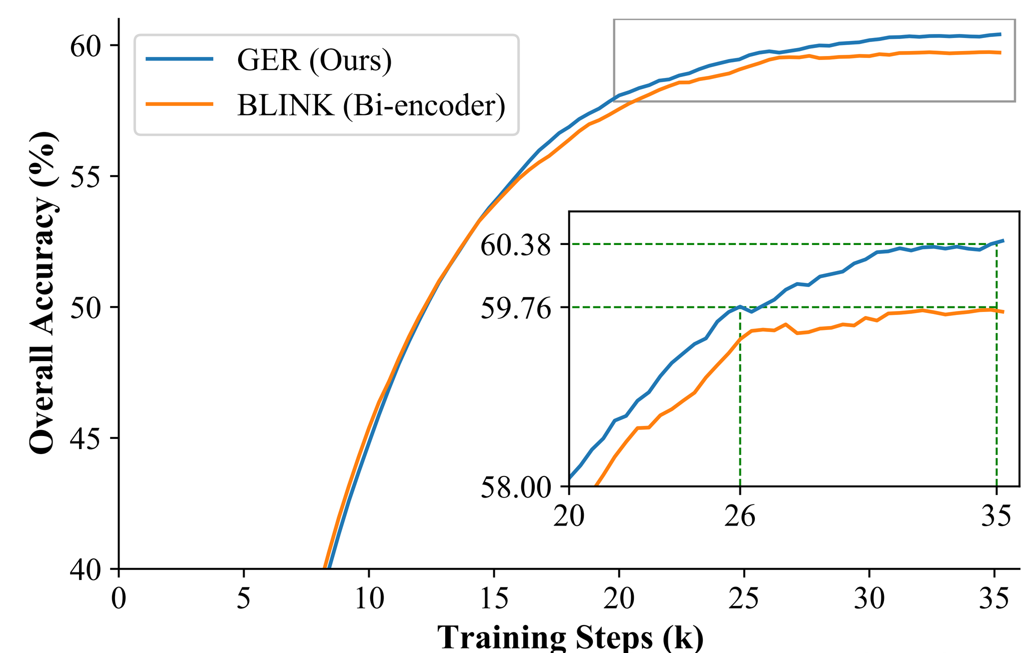
Mention Encoder	Entity Encoder	R@1	R@8	R@32	R@64
BERT	BERT	38.01	69.19	80.03	83.98
BERT+HGAN	BERT	38.16	69.41	80.04	83.92
BERT	BERT+HGAN	39.18	68.56	78.70	82.65
BERT+HGAN	BERT+HGAN	42.86	73.00	82.15	85.65

→ This table indicates that fusing fine-grained information for the encoder on one side will bring slight drops compared to the baseline BLINK. Meanwhile, the GER framework performs better than BLINK. For zero-shot entity retrieval, the mentions and entities should be embedded in one semantic space to match, thus fusing fine-grained information at one encoder leads to a drop in performance.

Overall Performance

To evaluate the overall performance, we employ the BERT-base based cross-encoder for entity ranking after retrieving 64 candidate entities.

Figure: Comparison of overall accuracy for BLINK and GER.



→ In conclusion, GER achieves a higher recall@64 than BLINK during entity retrieval stage with slight extra 8 minutes, and such higher recall@64 can save 2 hours in the entity ranking stage.

Resources

Email: wtq20@mails.tsinghua.edu.cn
Paper: https://arxiv.org/abs/2211.10991
Code: https://github.com/wutaiqiang/GER-WSDM2023